

Subject Selection in Applied Behavior Analysis

Andrew L. Homer

Missouri Department of Mental Health
and University of Missouri-Columbia

Lizette Peterson and Stephen A. Wonderlich
University of Missouri-Columbia

Past researchers have commented on the role of specifying relevant subject characteristics in determining the generality of experimental findings. Knowledge of subject selection criteria is important in interpreting and replicating research results. Such knowledge, as compared with many other historical and demographic characteristics of the subject, is likely to be related to a procedure's effectiveness. Data indicated that the majority of articles published in the *Journal of Applied Behavior Analysis* do not provide an adequate description of subject selection criteria. The failure to provide detailed information concerning subject selection criteria can prevent systematic replication of research results. The relatively low cost inclusion of complete descriptions of subject selection criteria would enhance the generality of applied behavior analysis research by facilitating systematic inductive manipulations and replications.

The science of applied behavior analysis relies on successive replications to determine the generality of research findings (Homer & Peterson, 1980). Unlike the more traditional experimental tack of assuming that results are generalizable to a wider population because of random subject selection and assignment to groups, behavior analysis uses cumulative corresponding and conflicting findings on single individuals to describe the parameters of a given procedure's utility (Hersen & Barlow, 1976). Cases in which findings agree extend the researcher's confidence in a technique, and cases in which the findings conflict serve as a spur for future research (Sidman, 1960). Tracking down the sources of such conflicting results is the primary method of determining the generality of any procedure (Kazdin, 1973; Leitenberg, 1973; Skinner, 1966).

SUBJECT VARIABILITY AND GENERALITY

One of the most important sources of

variability is the unique behavioral history of the subject (Sidman, 1960). Hersen and Barlow (1976) describe the importance of *subject generality*, the degree to which a procedure tested with one subject is applicable to another subject. They argue that evaluation of systematic replication is in part dependent upon the reader's ability to determine the important similarities and differences in the subjects treated. In each case in which subjects differ on relevant characteristics and yet respond similarly to the experimental manipulation, the generality of that procedure is enhanced. However, if researchers fail to describe the relevant characteristics of the subject, it is not possible to determine the extent to which previous findings are replicated, extended, or contradicted.

It is possible to specify, at least in theory, the relevant subject information which would best enhance the possibility of replication and allow judgment of the generality of the subject selected. A comprehensive behavioral history of the subject, complete with a functional analysis of all possible stimulus situations that could be encountered, would allow the reader to clearly judge the extent to which the subject in the experiment differed from other subjects that might be treated in a similar manner. Because all responses and all stimulus situations would be

This research was supported by a Summer Research Fellowship to the second author by the University of Missouri-Columbia Research Council. Reprints can be obtained from Lizette Peterson, Psychology Department, 210 McAlester Hall, University of Missouri-Columbia, Columbia, MO 65211.

represented, the subject's unique response to various cues, reinforcers, and punishers would be clear, and there would be no need to select relevant subject characteristics a priori. The characteristics that were relevant to the experimental effects would be apparent, after the fact. However, a complete behavioral history is possible only in theory. The expense and difficulty of producing the data, as well as the infeasibility of publishing data which to a large extent may be irrelevant, is prohibitive. How, then, can the important differences in subjects be determined and reported?

Relevant factors could be determined after the fact by continued manipulation of subject factors and replication. However, if relevant subject variables are not routinely reported, even post hoc manipulation of the unspecified critical variables might not be possible. There may be a number of ways in which the relevance of subject variables could be determined. A casual perusal of journals within applied behavior analysis will demonstrate that certain variables have typically been reported routinely. When human subjects are utilized, the gender, age, and any existing intellectual limitations of the subjects are almost always presented. Often, other demographic information such as diagnostic categories are also presented. This information is best categorized as subject description. The presentation of these descriptors does not usually reflect a systematic attempt to replicate findings across broad subject classifications. The usual descriptors presented may not even represent the experimenter's best guess about which subject factors may be relevant to the results. They are probably reported because of tradition, rather than a demonstrated or even suspected functional relation to the control of behavior. The presentation of subject descriptions, even those descriptors sanctioned by tradition, does not directly address the issue of developing generality through systematic replication. The question remains, which of the myriad possible subject descriptors, from average blood pressure to social history, might be relevant to the outcome of the

research? Are there practical methods of limiting the number of variables to be reported, yet maximizing the likelihood that potentially important variables will not be overlooked?

SUBJECT SELECTION CRITERIA

It is likely that increased knowledge concerning the influence of specific subject characteristics upon the efficacy of select procedures will progressively yield better answers to the question of which subject characteristics should be delineated in an experimental report. However, there is one source of information about subjects which might be of general importance to attempts to replicate experimental results. The criteria used by the experimenter to select the subjects represent specific and potentially important information about the subjects. If the subjects were not selected randomly, then the experimenter used some selection criteria. Because the experimenter is likely to be highly motivated to demonstrate experimental effects, it is likely that when selection criteria are used, they are believed by the experimenter to be relevant to the procedures to be studied. In comparison to a comprehensive behavioral history which may be infeasible because of a very high cost/benefit ratio, specifying the subject selection criteria would have a very modest cost (an extra few lines of journal space) and would have the potential benefit of yielding important information on characteristics of the subjects that might be relevant to the experimental effects reported.

In fact, McNamara and MacDonough (1972) argued that information should be presented on the criteria used to select subjects. Similarly, Sidman (1960) noted that subject selection criteria may play an important role in determining research results. Indeed, technological adequacy demands that enough information be presented to afford a typically trained reader a chance to replicate the results in a similar situation (Baer, Wolf, & Risley, 1968), and this information would necessarily include the criterion used to select the subject employed in the experi-

ment. Information on subject selection criteria is likely to have general importance regardless of the experimental design being employed. Its importance may be most clearly evident, however, in experiments which compare alternative treatments.

Subject Selection Criteria and Comparisons Between Procedures

Subject selection can directly be shown to affect conclusions when comparisons are made between different treatments. For example, two experiments that compared time out with positive practice could produce completely contradictory results if one study selected subjects who were particularly responsive to social contacts like hugs and verbal reinforcers, while the other study selected subjects who would not tolerate physical contact. With socially responsive subjects, time out from social reinforcers might be much more effective at decreasing responding than would a "hands-on" positive practice approach. With subjects who dislike social contact on the other hand, positive practice might be very effective, while time out might actually cause responding to increase. Because the number of comparative studies is increasing (Hayes, Rincover, & Solnick, 1980), such issues may become increasingly problematic.

Knowledge of subject selection criteria is thus important to interpreting results. The specification of subject selection characteristics also seems to be a sensible compromise between the routine reporting of comprehensive behavioral histories and the reporting only of routine demographic characteristics. However, it is unclear whether the majority of articles in applied behavior analysis currently include information on criteria used to select subjects.

Rate of Reporting Subject Selection Criteria

Because information on the criteria used to select subjects would appear to be a valuable tool in the pursuit of generality through successive replication and because this information would be vital in any case in performing an accurate

replication, an estimate was obtained of how often subject selection information was presented in one applied behavior analysis journal.

Two independent observers categorized the descriptions of subject selection criteria in the articles in the *Journal of Applied Behavior Analysis (JABA)* from 1968 to 1980. The primary observer rated every issue and the other observer rated one issue (25%) per year. The two observers substantially agreed with one another on the ratings (Cohen's 1960 κ ; $\kappa = .82$). Volumes were rated in random order. Only experimental articles were rated; "experimental" was arbitrarily defined as an article longer than three pages of text which included a method section. This definition excluded brief reports which might have sacrificed detailed subject selection information because of the condensed format as well as technical notes and theoretical presentations.

The survey revealed that 47.8%¹ of the articles published in *JABA* did not present any rationale for why the subjects were selected. In these articles, subjects were most often simply described, (e.g., "Mike was a 15 year old educable mentally retarded boy . . ."), and brief case histories were sometimes presented but no description of the reason for selecting these subjects was offered.

An additional 34.8% of the articles were judged as presenting an incomplete rationale. Articles in this category presented a selection rationale but the description of the rationale was not exhaustive, and thus would not permit replication. These articles typically selected subjects because of high or low rates of particular behaviors (e.g., "The subject was one of the least vocal," "Karen was selected because she did not use the plural form"). Articles in this category also included referrals, such as, "Bill was referred to the center by the school district." In all cases, some infor-

¹ The percentages sum to more than 100% because articles reported more than one experiment or used more than one set of subjects.

mation on selection criteria was given, but it was not clear how many other subjects with the same characteristics were rejected or whether or not other, unspecified criteria influenced subject selection.

Finally, 23.1% of the articles presented a complete rationale for subject selection. Articles in this category included experiments using all available subjects, randomly selecting subjects from some specified group, or providing detailed descriptions of the selection procedure.

The data demonstrate that while some descriptions of subject characteristics are presented in the majority of articles in *JABA*, only a small percentage of articles reported on the complete criteria used to select the subject. In spite of past methodological comment on the importance of subject characteristics and subject selection (e.g., Hersen & Barlow, 1976; McNamara & MacDonough, 1972; Sidman, 1960) and the low cost of such reporting, the complete reporting of subject selection criteria is the exception rather than the rule. Thus, concern about the relative absence of this potentially important information would seem to be warranted.

REASONS FOR THE FAILURE TO REPORT SUBJECT SELECTION CRITERIA

There are a variety of possible reasons why subject selection criteria are not routinely reported. First, applied behavior analysis has its roots in animal experimentation, and the selection of subjects (at least prior to experimental manipulation) is less problematic there because of the high level of experimental control and homogeneity of the subject pool.

Second, selection criteria may not be based on data the experimenter possesses, and the experimenter may be reluctant to speculate on processes that are not demonstrated empirically. Thus, Susie, a mute 8-year-old, may be selected because she seemed to be very responsive to social stimuli and the technique to be used relies on social rewards. Because the experimenter has only anecdotal data on Susie's social responsivity, however, her

demographic characteristics are presented but the criteria used to select her from other mute 8-year-olds is not. Thus, the variable that might be the most important in systematic replication is not even mentioned. Similarly, the experimenter may select Johnny simply because Johnny is very aggressive or because Johnny is both aggressive and susceptible to peer pressure. If the experimenter reports that Johnny was randomly selected from the six most aggressive boys in the class, the reader is assured that there was no second criterion. If the experimenter uses only the label "aggressive," no such assurance is provided. It would, of course, be useful to have a functional analysis of behavior to document that Susie was unusually sensitive to social stimuli or that Johnny was more aggressive than other boys in his class. However, requiring extensive documentation will raise the cost/benefit ratio to one which is likely to be intolerable to researchers. Where this is not the case, such analyses should be included. It is suggested here that, at minimum, the selection criteria be reported along with whatever data the experimenter used to establish the criteria. If the data are intuitive or anecdotal, better to report that than remain with the status quo of reporting no criteria or incomplete criteria. Obviously, data are better than opinions and demonstrations of functional relationships are better than intuition, but even opinions and intuition are better than no information, especially when the absence of information can be misleading. Opinions can be explicitly analyzed with future data, and thus are better than nothing at all.

Failing to report the data out of custom or because one lacks quantifiable data to demonstrate the criterion actually blocks the possibility of obtaining quantified data later on. It is of vital importance to note conditions in which a procedure fails and to separate those conditions experimentally from those in which the technique succeeds (Hersen & Barlow, 1976). Yet failures to replicate are rarely published (Homer & Peterson, 1980), and if the reason for the failure to replicate is a difference in subject selection criteria

not reported by the successful and unsuccessful investigators, there is no possibility for systematic analysis of the characteristic. The tack urged by Sidman (1960) of systematically manipulating the relevant subject variables to isolate the parameters of the technique is possible only if one can identify these variables.

Finally, experimenters might not list the selection criteria because they fear such a listing may imply that their experimental procedure is limited to subjects with that particular characteristic. Thus, they may fear in the example above that the reviewer might argue that "of course, the procedure worked with a socially responsive child, but this isn't a very robust test." While this is a possibility, it would be an error on the part of the reviewer, not the experimenter. Sidman (1960) noted that one never loses generality in data by limiting the population to which a given experimental result applies.

In point of fact, generality is increased. It is unrealistic to expect that a given variable will have the same effects upon all subjects under all conditions: As we identify and control a greater number of the conditions that determine the effects of a given experimental operation, in effect we decrease the variability that may be expected as a consequence of the operation. It then becomes possible to produce the same results in a greater number of subjects (p. 190).

Thus, failing to report a selection criterion because one fears it implies limited generality actually functions to limit generality.

COSTS AND BENEFITS OF REPORTING SUBJECT SELECTION CRITERIA

The tradition in applied behavior analysis has therefore been to fail to include subject selection criteria. It is perhaps not sufficient to call for a change in scientific practice based on logic alone. Changing the way studies are reported also requires consideration of the costs involved in the present and proposed practice and of the possible benefits. Unfortunately, the very nature of the present practice of not fully describing subject selection criteria precludes assessment of its effects. The necessary data are simply not available to assess the influence of ex-

perimenter selection practices. The costs of the present practice are, therefore, unknown, as are the potential benefits of the proposed change. The costs to researchers and journals of changing the practice to include complete specification of potentially relevant past behaviors may be prohibitive. Where cost is not prohibitive, such data should certainly be presented. However, it is argued here that the cost of a compromise, including subject selection criteria with routine demographic characteristics, is trivial. This small cost of altering the selection description surely outweighs the potential cost of the present practice in which subject characteristics that the experimenter believes may be important to experimental effects are not even mentioned.

Of course, instigating such a change, while of very low cost, may be of little corresponding benefit if experimenters do not currently systematically select subjects on the basis of certain characteristics. It would seem logical to assume that they do, however, simply because the experimenter is likely to be aware of subject differences and to be highly motivated to effect a successful outcome by selecting the subject whose characteristics make him or her most amenable to the manipulation used. Even if only a minority of subjects are presently selected on the basis of an unspecified criterion, the present method of reporting does not allow this aspect of subject characteristics to be evaluated. In other words, the present methodology leaves the reader wondering if the subject was selected because of certain characteristics, was selected randomly from a group of available subjects, or was the only such subject available to the experimenter. This is particularly frustrating for the individual failing to replicate a finding. Did the successful experimenter know or suspect something about his subject that the experimenter failing to replicate did not know? With the present practice, the answer simply is not clear.

TOWARD A SOLUTION

There are thus a number of problems possible when subject selection criteria are

not specified. The solution suggested here is simply the complete specification of the selection process, whether it is random, based on the limitations of the subject pool, or some specific subject characteristics. It is important to note that the solution advocated here is *not* the use of random subject selection based on the logic of statistical inference as developed by Fisher (1935) and espoused by Campbell and his colleagues (Campbell, 1957, 1963, 1969; Campbell & Stanley, 1966; Cook & Campbell, 1976, 1979). While this system is widely accepted, behavior analysis achieves generality in a different way, through successive replication (Herson & Barlow, 1976; Johnston & Penypacker, 1980; Sidman, 1960). Neither is the solution a progressive move toward encouraging experimenters to simply speculate about relevant subject characteristics. The authors have no desire to be associated with a position which eschews data in favor of an arm-chair psychology consisting of researchers' conjectures and opinions. It would certainly be consistent with the rhetoric of behavior analysis to call for complete behavioral histories, empirical demonstrations of suspected variables such as social responsivity, and demonstrations of the effectiveness of reinforcers. We seek a compromise between rhetoric and practicality.

The solution which is advanced here is consistent with the method of applied behavior analysis which demands that all subject selection criteria be fully described (e.g., Sidman, 1960). Anything less may prevent both direct and systematic replication and the use of inductive technique building manipulations which are the defining characteristics of behavior analysis. The logic of the analysis of behavior does not require random subject selection to obtain generality. It does, however, require systematic replication to obtain generality, and systematic replication can best occur when subject selection criteria are sufficiently explicated to permit direct and systematic replication.

Successful replication does not, of course, depend exclusively on control

over variation introduced by using different subjects. A variety of factors including subject characteristics, therapist characteristics, setting variables, and even the particular target behavior selected, could influence the outcome of the experiment. The strategy suggested here for dealing with subject variables could be generalized and applied to any factor that an experimenter explicitly selects for inclusion or exclusion. An explication of any selection criterion will assist in the assessment of the generality of the research. We suggest here as a beginning the inclusion of all relevant information concerning the choice of a specific subject for a given experimental protocol.

REFERENCES

- Baer, D. M., Wolf, M. M., & Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 91-97.
- Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 1957, 54, 297-312.
- Campbell, D. T. From description to experimentation: Interpreting trends as quasi-experiments. In E. W. Harris (Ed.), *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.
- Campbell, D. T. Reforms as experiments. *American Psychologist*, 1969, 24, 409-429.
- Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.
- Cohen, J. A. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20, 37-46.
- Cook, T. D., & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Skokie, Ill.: Rand McNally, 1976.
- Cook, T. D., & Campbell, D. T. *Quasi-Experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally, 1979.
- Fisher, R. A. *The design of experiments*. Edinburgh: Oliver & Boyd, 1935.
- Hayes, S. C., Rincover, A., & Solnick, J. V. The technical drift of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1980, 13, 275-285.
- Hersen, M., & Barlow, D. H. *Single case experimental designs*. New York: Pergamon Press, 1976.
- Homer, A. L., & Peterson, L. Differential reinforcement of other behavior: A preferred

- response elimination procedure. *Behavior Therapy*, 1980, 11, 449-471.
- Johnston, J. M., & Pennypacker, H. S. *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum, 1980.
- Kazdin, A. E. Methodological and assessment considerations in evaluating reinforcement programs in applied settings. *Journal of Applied Behavior Analysis*, 1973, 6, 517-531.
- Leitenberg, H. The use of single-case methodology in psychotherapy research. *Journal of Abnormal Psychology*, 1973, 82, 87-101.
- McNamara, J. R., & MacDonough, T. S. Some methodological considerations in the design and implementation of behavior therapy research. *Behavior Therapy*, 1972, 3, 361-378.
- Sidman, M. *Tactics of scientific research*. New York: Basic Books, 1960.
- Skinner, B. F. Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Crofts, 1966.